Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: www.jsoftcivil.com

Discovering Bengkulu Province Earthquake Clusters with CLARANS Methods

Arie Vatresia [1,2,*]; Ferzha Putra Utama [3]; Intan Permata Hati [4]; Lindung Zalbuin Mase [5]

1. Associate Professor, Department of Informatics, Faculty of Engineering, University of Bengkulu, Bengkulu, Indonesia
2. Center for Computing Research, National Research and Innovation Agency
3. Lecturer, Department of Information System, Faculty of Engineering, University of Bengkulu, Bengkulu, Indonesia
4. Graduate Student, Department of Informatics, Faculty of Engineering, University of Bengkulu, Bengkulu, Indonesia
5. Associate Professor, Department of Civil Engineering, Faculty of Engineering, University of Bengkulu, Bengkulu, Indonesia

* Corresponding author: *arie.vatresia@unib.ac.id*

https://doi.org/10.22115/SCCE.2023.381204.1589

ARTICLE INFO

Article history:
Received: 13 January 2023
Revised: 05 May 2023
Accepted: 31 December 2023

Keywords:
Earthquake;
Bengkulu;
Clustering;
Elbow;
Clarans.

ABSTRACT

Bengkulu is one of the provinces that lies in the ring of fire which has dynamic number of earthquakes each year. There were 2,989 earthquakes recorded in the Bengkulu from 2000 to 2021 with spatially different and unique attributes. The occurrences over Bengkulu were also included on the top twenty of significant event of earthquake that caused a lot of damage and death with the highest level of magnitude (8.4 SR). Due to its novel occurrences, how the points of earthquake clustered is still an interesting research question to be answered to empowered the location to be ready for the disaster. With the accelerating movement of artificial intelligence, this research used three methods to process earthquake data (Elbow, CLARANS, and Silhouette Coefficient) to find a better understanding of the cluster's discovery. Matuschka method was occupied to uncover the earthquakes intensity and level of damage caused by earthquakes in the areas based on the MMI scale. This research succeeded to generate the six best cluster types based on the time of occurrence and level of damage during with the information. Furthermore, this research also mapped the severity of risk over the district to see the distribution of clusters produced. This research found that the area of Rejang Lebong is the most vulnerable to the earthquake with the class of heavy damage based on MMI value. Another class that involved into moderate risk is Muko-Muko area. The Validation performance showed a value of 0.55 which involved on feasible rate.

How to cite this article: Vatresia A, Utama FP, Hati IP, Mase LZ. Discovering bengkulu province earthquake clusters with CLARANS methods. J Soft Comput Civ Eng 2024;8(3):71–86. https://doi.org/10.22115/scce.2023.381204.1589

2588-2872/ © 2024 The Authors. Published by Pouyan Press.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Indonesia is a maritime country located at the confluence of three tectonic plates: the Eurasian plate, the Indo-Australian plate, and the Pacific plate [1–5]. These conditions make Indonesia vulnerable to volcanic eruptions, earthquakes, tsunamis, floods, and landslides [6–8]. In addition, Bengkulu Province is also located between two active earthquake fault lines: the Mentawai fault epicenter and the Sumatra fault epicenter around Bukit Barisan. These conditions make Bengkulu Province prone to earthquakes [7–9]. Mase et al. [8] mentioned that within the last two decades, two strong earthquakes with a magnitude of more than 7 happened in Bengkulu and resulted in massive damage. Those earthquakes were known as the Bengkulu-Enggano Earthquake in 2000 and the Bengkulu-Mentawai Earthquake in 2007. Farid and Mase [7] also highlighted that, based on damage intensity level prediction using Modified Mercalli Intensities (MMI), both earthquakes could result in a damage level up to Scale IX in Bengkulu. It indicates that Bengkulu is very vulnerable to experiencing seismic damage. In line with this, Misliniyati et al. [10] and Mase [11] suggested that facing the reality of seismic vulnerable high-risk, several studies in seismic hazard mitigation should be performed to reduce the potential seismic damage in Bengkulu City. Those studies can be initiated by predicting damage intensity level and site vulnerability, as conducted by Sugianto et al. [12] and Refrizon and Mase [13].

Artificial intelligence (AI) is increasingly used in various fields, including seismology [14,15]. One branch of artificial intelligence technology is machine learning [16,17]. Machine learning (ML) is learning computer algorithms to find patterns or make predictions without human involvement [16]. Machine learning is broken down into two categories: supervised learning and unsupervised learning [17]. In this study, we used unsupervised (clustering) machine learning technology. Clustering is a process of grouping data based on the similarity of features or characteristics [18]. Farid and Mase said that Bengkulu is an area prone to earthquakes. According to data published by the National Agency for Meteorology, Climatology, and Geophysics, or locally named the *Badan Meteorologi, Klimatologi, dan Geofisika* (BMKG), there were 2,989 earthquakes in the Bengkulu Province from 2000 to 2021. Seeing the number of earthquake events that have occurred, artificial intelligence is one way to process earthquake data to determine the severity level in Bengkulu Province. The data used to determine the severity of an earthquake are latitude, longitude, depth, strength and date. In this case, data management would be carried out to produce earthquake data clustering and generate earthquake damage intensity levels based on the Peak Ground Acceleration (PGA) value. PGA is the main parameter that could reflect the damaging impact. Farid and Mase [7] and Mase [8] also noted that PGA does not only reflect the potential damage but also the maximum potential forces for structural building design. The large number of earthquakes in Bengkulu makes this research necessary to determine the pattern of distribution and grouping. There is still a lack of research discussing earthquake clustering, especially in the Bengkulu Province, making this research important to do. The grouping of earthquakes is based on the location of the epicenter, the depth of the epicenter, the time of occurrence, the depth, and the strength of the earthquake. Clustering is the grouping of data into several similar classes [19]. The clustering process in this research consisted of three stages: pre-cluster analysis, cluster analysis, and cluster validation [2,20,21] The first stage used the Elbow method to analyze the pre-cluster. In the second stage, cluster analysis used the

CLARANS method, and the last stage was cluster validation using the silhouette coefficient method. The use of the clustering method in determining the PGA values could be applicable. In addition, the use of PGA is also applicable to predict the damage intensity level in the study area. Therefore, it is important to implement clustering analysis to estimate the tendency of earthquake and the possible impact that could result.

Earthquake disasters have always been identified by the damage they cause. This relates to PGA and MMI values. These two things significantly affect the clustering of earthquakes in Bengkulu later. PGA is the most significant ground vibration acceleration value ever recorded in an earthquake-affected location [4,22–24] PGA is a critical factor affecting an earthquake's intensity [4,22] In addition to the acceleration of vibrations, the earthquake's strength also creates damage. Earthquake strength is classified based on the level of damage inflicted by an earthquake in a particular place. MMI scales are commonly used to measure the extent of damage from earthquakes. The MMI scales are a unit for measuring the magnitude of an earthquake [25].

2. Research significance

This study demonstrates a new approach to data processing using three methods to cluster earthquake data [26–28]. In addition, this study can also classify the damage level to buildings into five different levels based on the value of the damage scales (MMI)[29]. The system built in this research is WebGIS. Therefore, the general public can find data on earthquakes in Bengkulu Province based on the clustering results and determine the damage level to buildings in each area of Bengkulu Province. Thus, can be benefit for further recommendation of infrastructure development that still on the list of government in Indonesia.

3. Methods

The system development method used in this research is the Cross Industry for Data Mining (CRISP-DM) method. CRISP-DM is a hierarchical process model consisting of six stages (Fig. 1): business understanding, data understanding, data preparation, modeling, evaluation, and deployment [30–32].

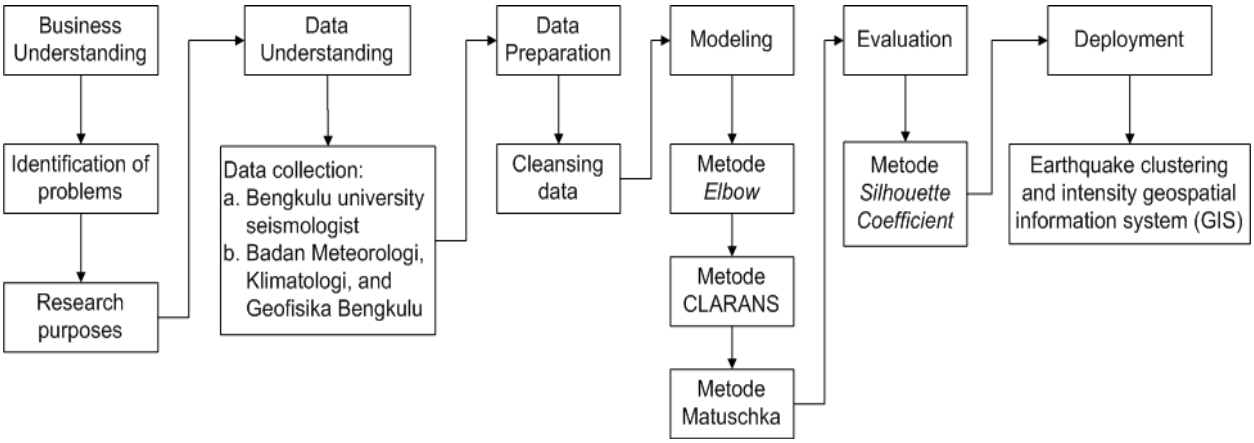


Fig. 1. CRISP-DM Flow.

The business understanding stage explains research problems, research objectives, and types of data mining [33]. The problem in this study is that Bengkulu Province is prone to earthquakes. During the last 21 years, 2,989 earthquakes have been recorded [25]. In this case, data management will be carried out to produce earthquake data clustering and classify earthquake intensity values based on the PGA value, so that it can help the people of Bengkulu know the data on earthquakes that have occurred in Bengkulu Province based on the results of clustering. Also, they can find out the level of damage to buildings experienced by each area, which is calculated from the scale of the damage. The data understanding stage is carried out by collecting, exploring, and understanding the data [30,32,33] The data used is tectonic earthquake data. Shear fractures caused by tectonic earthquakes cause ground crack [34]. The data for this study came from two sources: (a) Mr. Lindung Zalbuin Mase, an earthquake expert at Universitas Bengkulu, from 2000 to 2009 [7]; and (b) the *Badan Meteorologi, Klimatologi, and Geofisika Pelayanan Terpadu Satu Pintu* BMKG Bengkulu from 2010 to 2021. The attributes of earthquake data are date, latitude, longitude, depth, and magnitude. The magnitude of the earthquake used is greater than or equal to 3. Data Preparation stage involved the process of selecting data and the process of cleaning up unused data are described [35,36]. The data preparation stage to fix problems in the data and ensure that the data is correct for use [30,32,33]. At the data preparation stage, there are two processes: cleaning and changing the data format. The cleaning process (a) removes the time attribute; (b) converts date data format to number format. Because the date attribute will be used in the clustering process, numerical values (dates) are stored as sequential serial numbers to be used in the calculation process. By default, January 1, 1900, is serial number one.

The modeling stage begins with the selection of a modeling technique [32,33,35]. The research consists of three stages: pre-cluster analysis, cluster analysis, and earthquake intensity. The pre-cluster analysis uses the Elbow method. The Elbow method is one method for determining the optimal number of clusters in data by comparing the sum of square error values that will form an Elbow at a point [20,21,37–39]. The cluster analysis uses the CLARANS method. The CLARANS method can produce more effective results than the PAM and CLARA algorithms [19,40]. The CLARANS method is known as an "alternative" because, at the time of the medoid iteration, the medoid point with the smallest number of distances from other cluster members (minimum cost) is chosen. The earthquake intensity uses the Matuschka method [41]. Peak ground acceleration (PGA) is a critical factor influencing earthquake intensity determination [42,43]. The equation (1) used from Matuschka (1980) to see the rate of PGA.

$$\mathrm{a} = 119.e^{0.81\,\mathrm{R}}.(\mathrm{H}+25)^{-1.15} \quad (1)$$

Where a is PGA (cm/sec), e is the natural number 2,718, R is the earthquake strength (SR), and H is the hypocenter (km). The hypocenter value is obtained from a calculation that requires the epicenter variable and the earthquake's depth. The latitude and longitude variables are used to calculate the epicenter value.

At the evaluation stage, the results are checked against the defined business objectives [33,38,44]. In this research cluster validation by using the silhouette coefficient method. The Silhouette Coefficient method determines each cluster's quality by calculating each cluster's

average silhouette index value [21,45,46]. The deployment stage is described generally in the user guide. It could be a final report or a system [30,33,35,36]. The system consists of two parts: computing and information. In the computing section, the system will process earthquake data. The results of earthquake data clustering are presented in the information section. WebGIS will be used to display the results of the earthquake intensity.

4. Results and discussion

4.1. Clustering

Bengkulu Province is one of the provinces that lies in the ring of fire. It means the area around the epicenter felt the impact of the earthquake. In the case of clustering earthquakes, it is carried out in several stages, as shown in Fig. 2. Currently, the application of computer science in the field of civil engineering is used. The grouping of the degree of vulnerability to disaster effects, as done in this study, has been done before [47]. Labeling on each cluster commonly done in the field of computer science is also used in civil science for clustering [48].

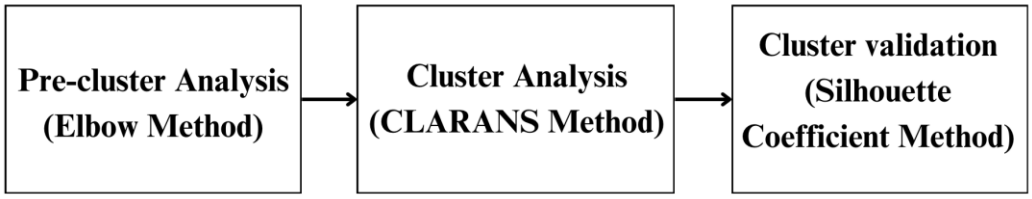


Fig. 2. Clustering stages.

The first stage in the clustering process is pre-cluster analysis. A pre-cluster analysis is a stage to determine the number of clusters or K values to be used. BMKG clusters earthquakes based on the depth of the earthquake point and produces three clusters: shallow, middle, and deep. Identification of the number of clusters or the value of K (cluster numbers) is the most critical and primary step in the clustering process [2]. If the number of clusters is determined arbitrarily (unsuitable), then the cluster results will not be maximized. To overcome this, we need a method to determine the number of clusters or the value of K. Elbow method produces optimal clusters in the clustering optimization process. This research will apply the Elbow method to determine the optimal number of clusters used in CLARANS clustering. The first step in the Elbow method is to calculate the sum of squared errors (SSE) with equation (2).

$$SSE = \sum_{k=1}^{k} [||latitude_x - latitude_y||]^2 + \left[\left|longitude_x - longitude_y\right|\right]^2 + \left[\left|depth_x - depth_y\right|\right]^2 + \left[\left|strength_x - strength_y\right|\right]^2 \quad (2)$$

The second step looks at the results of the sum of square error (SSE) values that have the largest difference or a significant decrease with equation (3).

$$Difference = SSE_i - SSE_{i+1} \quad (3)$$

In this study, K = 6 was found to be the optimal K value. It can be seen from the SSE values that form the elbow in Fig.3 or the most significant difference in SSE values in Table 1.

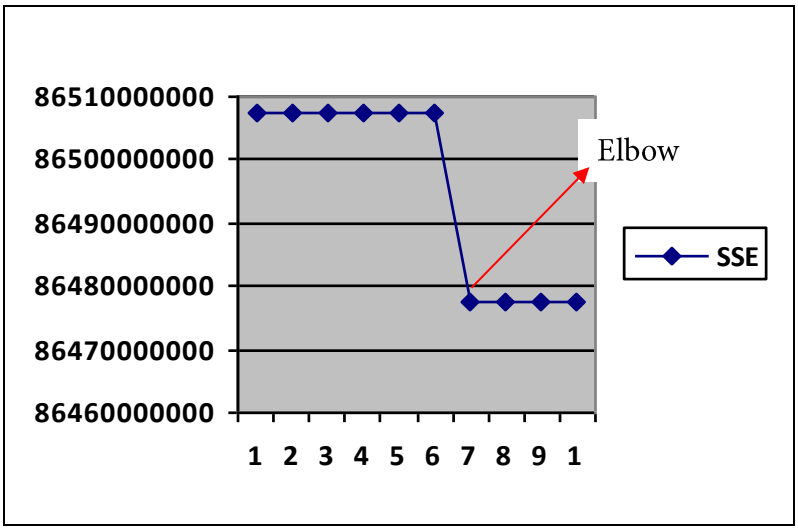


Fig. 3. Elbow calculation result graph.

Table 1
Sum of square error result.

Cluster	SSE Result	Difference
1	86507394530.003	0
2	86507385391.133	9138.870010376
3	86507383984.054	1407.078994751
4	86507380692.854	3291.1999969482
5	86507379998.717	694.13700866699
6	86477638171.292	29741827.424988
7	86477638111.79	59.502014160156
8	86477637259.418	852.37199401855
9	86477637012.046	247.37199401855
10	86477636984.179	27.86700439453

After the value of K or the number of clusters is determined, the next step is cluster analysis. Cluster analysis is a statistical method that does not have assumptions or prior knowledge about the data. Until the data is grouped only based on the similarity of the data [45,49], extraction of information from big spatial data can use spatial clustering with the CLARANS method. CLARANS analyzes spatial and non-spatial attributes [50]. This study will cluster earthquakes in Bengkulu Province using CLARANS on five earthquake data attributes: two spatial (latitude and longitude) and three non-spatial (magnitude, depth, and date). The results of earthquake data clustering in Bengkulu Province for 2000-2021 can be seen in Table 2.

Table 2
Clustering Result from CLARANS.

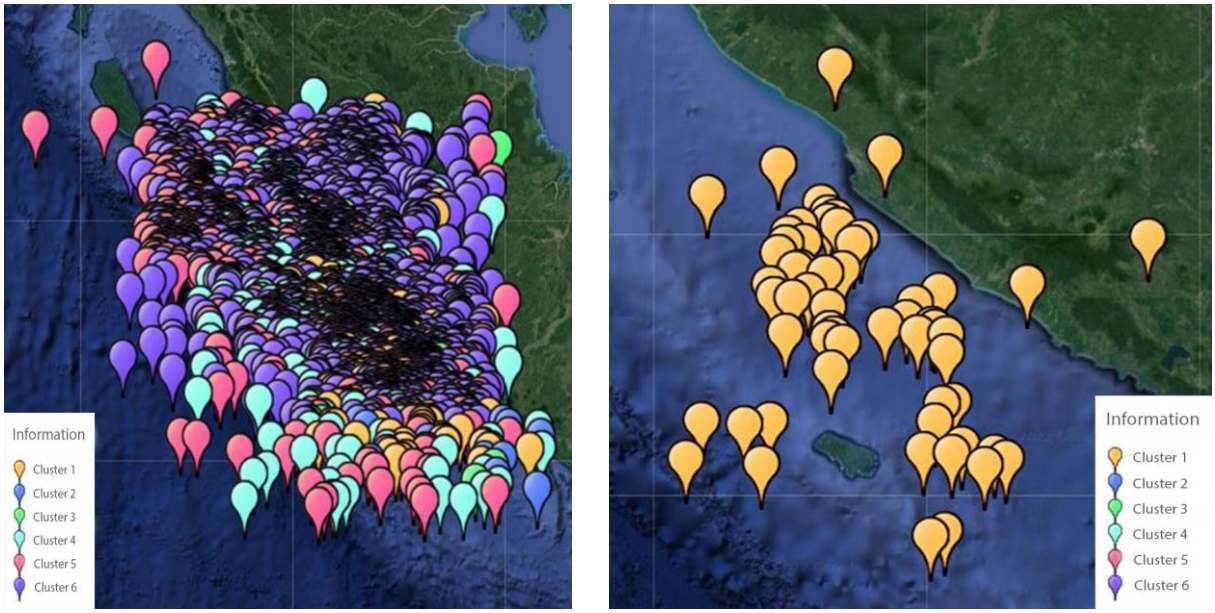
Cluster	Latitude	Longitude	Magnitude (SR)	Depth (KM)	Date	Data
1	(-6.13)-(-1.99)	(100.09)-(104.39)	(4)-(8)	(28)-(219)	(14-03-2000)-(16-06-2002)	187
2	(-6.41)-(-2.73)	(101.2)-(104.53)	(5)-(6)	(10)-(98)	(02-10-2002)-(24-10-2004)	31
3	(-6.33)-(-2.43)	(100.62)-(103.98)	(4.4)-(7.3)	(25)-(582)	(23-05-2004)-(06-04-2008)	19
4	(-6.55)-(-1.81)	(99.99)-(104.54)	(4)-(8.5)	(10)-(103)	(06-04-2005)-(17-01-2010)	269
5	(-6.61)-(-1.39)	(97.84)-(104.11)	(3.1)-(7.2)	(3)-(132)	(03-05-2009)-(26-01-2018)	467
6	(-5.66)-(-2.06)	(99)-(103.89)	(3)-(8)	(3)-(750)	(01-02-2011)-(30-12-2021)	2016

The CLARANS method successfully clustered data into six groups. If seen (Table 2) the data is clustered for a time of occurence. The variables latitude, longitude, magnitude, and depth do not greatly affect this clustering. Data that has been clustered must be interpreted. In this case, the interpretation process uses a decision tree. A decision tree is a structure like a flowchart for naming labels or describing data [50–52]. With the interpretation process using the decision tree, it is very clear that the earthquake data for Bengkulu province from 2000-2021 are clustered based on the time of occurrence (Table 3). In the decision tree, there is a Gini coefficient. Gini is the standard coefficient applied to hierarchical clustering to determine whether the sum of the data in all clusters participates well [53,54]. The Gini coefficient ranges from 0–1; closer to 0 indicates that the grouping is more even. And conversely, closer to 1 indicates the grouping is uneven [51,53,54].

Table 3
Interpretation using a decision tree

Cluster	Criteria
C1	187 data clustered by date 09-08-2002 with a gini of 0.349. So cluster 1 consists of earthquake data that occurred from 13-03-2000 to 09-08-2002.
C2	30 data clustered by date (09-08-2002 > cluster 2 ≤ 2 04-05-2004) with a gini of 0.465. 1 data clustered by date (09-08-2002 > cluster 2 ≤ 04-05-2004) and earthquake magnitude ≤ 17.5 SR with a gini of 0.1.
C3	18 data clustered by date (04-05-2004 > cluster 3 ≤ 26-03-2005) with a gini of 0.465 and earthquake magnitude 17.5 SR with a gini of 0.1. 1 data clustered by date (06-03-2008 > cluster 3 ≤ 15-02-2009) with a gini 0.029.
C4	65 data clustered by date (26-03-2005 > cluster 4 ≤ 15-02-2009) a gini of 0.576 and latitude ≤ (-4.58) with a gini of 0.29. 2 data clustered by date (26-03-2005 > cluster 4 ≤ 15-02-2009) with a gini 0.044 and latitude (-4.58) with a gini of 0.29. 201 data clustered by date (26-03-2005 > cluster 4 ≤ 15-02-2009) with a gini 0.576 and latitude > (-4.55) with a gini of 0.007. 1 data clustered by date (15-02-2009 > cluster 5 ≤ 31-01-2010) with a gini 0.05 and latitude > (-1.895) with a gini of 0.004.
C5	465 data clustered by date (15-02-2009 > cluster 5 ≤ 31-01-2010) with a gini of 0.6556 and latitude ≤ (-1.895) with a gini of 0.004. 1 data clustered by date (31-01-2010 > cluster 5 ≤ 31-01-2011) with a gini 0.5 and latitude (-1.895) with a gini of 0.004. 1 data clustered by date > (31-01-2011) with a gini 0.509, longitude > 102.715 with a gini of 0.005, and earthquake magnitude ≤ 3.75 SR with a gini of 0.278.
C6	1586 data clustered by date > (31-01-2011) with a gini 0.509 and longitude ≤ 102.705 with a gini of 0.001 5 data clustered by date > (31-01-2011) with a gini 0.509 and longitude > 102.715 with a gini of 0.005, and earthquake magnitude > 3.75 SR with a gini of 0.278. 425 data clustered by date > (31-01-2011) with a gini 0.509 and longitude > 102.715 with a gini of 0.001.

The results of earthquake data clustering will be implemented in WebGIS. The clustering results are displayed as a map on the earthquake data clustering page (Fig. 4(a)). The data will be in the form of coordinate points based on clusters. Several selected categories can be used on this page to display the clustering results. The categories are: (1) Date: This date category will display the data based on the selected date or time range; (2) Clustering: This clustering category will display the selected cluster data. Fig. 4(b) shows the date and cluster functions in action. Fig. 4(b) shows the results of clustering earthquake point data from 01-01-2000 to 31-12-2000, with only Cluster 1 (C1) included.



a b

Fig. 4. (a) Earthquake data clustering page, (b) Results clustering (01-01-2000)-(01-01-2001).

The last step of clustering is cluster validation. The method used in determining the quality of each cluster formed is the silhouette coefficient method. The Silhouette Coefficient method is a method for determining each cluster's quality by calculating the silhouette index's average value. The silhouette coefficient value of cluster 1 is 0.75; the silhouette coefficient value of cluster 2 is 0.52; the silhouette coefficient value of cluster 3 is 0.42; the silhouette coefficient value of cluster 4 is 0.65; the silhouette coefficient value of cluster 5 is 0.66; and the silhouette coefficient value of cluster 6 is 0.28. The global or overall silhouette coefficient value is 0.55. According to Kaufman and Rousseeuw in 2005 silhouette coefficient value 0.51-0.70 is feasible cluster. So, the clustering process in this study can be said to be successful with a cluster validation value of 0.55 (feasible cluster).

4.2. Earthquake intensity based on cluster

Indonesia uses the MMI scales to express the intensity or impact caused by an earthquake [4,5,41,55,56]. Earthquake intensity is the level of damage caused by an earthquake at a specific location, expressed on the MMI scales. The earthquake's intensity needs to be calculated in each area, so the MMI scales differ in every area. Locations close to the epicenter have large MMI scales.

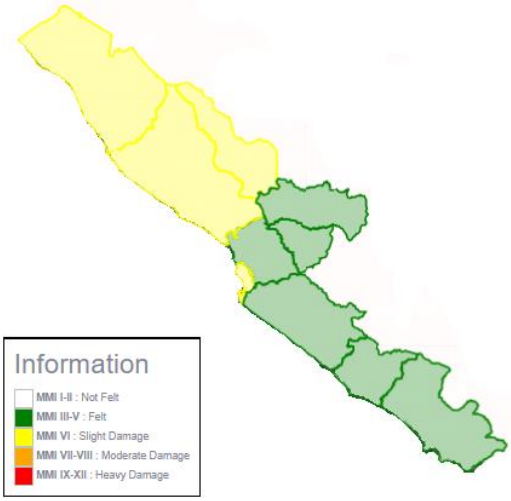
Table 4
MMI scales colour.

Color	Description	MMI Scales
White	Not Felt (Not felt or felt only by a few people but recorded by the device)	I-II
Green	Felt (Sensed by crowds but does no damage. However, the light objects that were hung swayed, and the glass windows shook)	III-V
Yellow	Slight Damage (Non-structural parts of the building suffered minor damage, such as hair cracks on the walls, tiles sliding down, and some falling)	VI
Orange	Moderate Damage (Many cracks occurred in the walls of simple buildings, some collapsed, broken glass. Some of the plaster of the walls came off. Most of the tiles slide down or fall off. The structure of the building suffered light to moderate damage)	VII-VIII
Red	Heavy Damage (Most of the permanent building walls collapsed. The structure of the building was heavily damaged, with curved railroad tracks)	IX-XII

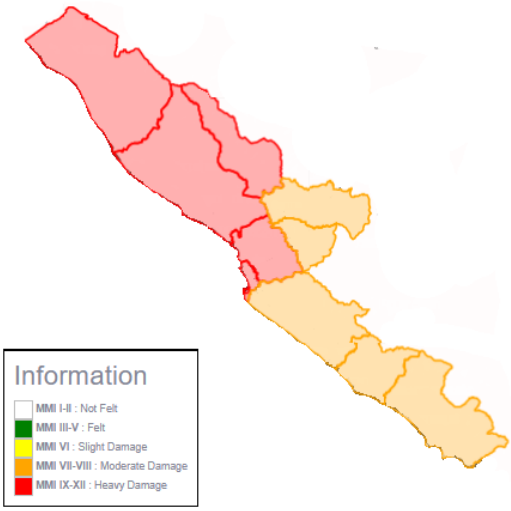
Calculating the value of earthquake intensity can be done using the following approaches: Matuschka, Donovan, and Midorikawa. The Matuschka method for obtaining the best (most accurate) damage prediction [41]. The first step in calculating the earthquake intensity value is determining the location coordinates (latitude and longitude) of each district in Bengkulu Province. The function of determining the point is to determine the value of the earthquake intensity in each district so that it can find out the level of damage to buildings experienced by each area, which is calculated from the scale of damage (MMI). The second step is to determine the epicenter values. The epicenter values represent the epicenter of the earthquake on the earth's surface [1,4,57]. The epicenter values are obtained by calculating the latitude and longitude values of the earthquake data against the latitude and longitude values of each district in Bengkulu Province. The hypocenter value is calculated after the epicenter value has been determined. The hypocenter of an earthquake is the location on Earth where it occurs [1,6,41]. The third step is calculating the PGA values. As a result, the most significant ground vibration acceleration values ever recorded in an earthquake-affected location are PGA [22,41,56]. After obtaining the PGA values, the next step is calculating the earthquake intensity values. Earthquake intensity is the level of damage caused by an earthquake at a specific location. The map depicts the intensity results' implementation (Fig. 5(a), 4(b), and 4(c)).

Fig. 5. (a) shows different options for displaying the level of building damage in each region of Bengkulu Province between 2000 and 2021.The categories are: (1) Choose the average MMI: selected average MMI category will display the selected MMI average data. The selection of this MMI average affects the damage level map of the building. (2) Year: This year's category will display data on the damage level to buildings in the selected year. In 2007, the PGA values and the most significant overall average MMI were recorded. As seen in Fig. 5. (b), Bengkulu City, Muko-Muko District, Bengkulu Utara District, and Lebong District suffered minor damage, marked in yellow. Buildings in the Rejang Lebong District, Kepahiang District, Bengkulu Tengah District, Seluma District, Bengkulu Selatan District, or Kaur District have not been damaged. When the average MMI value of eight is used and the year 2000 is chosen, only Bengkulu and Kepahiang districts are colored red (Fig. 5. (c)). According to the BMKG in 2022, the red color indicated heavy damage (Table 4).

(a)



(b)



(c)

Fig. 5. (a) Earthquake intensity page, (b) Average overall MMI in 2007, (c) Average MMI ≥ 8 in 2007.

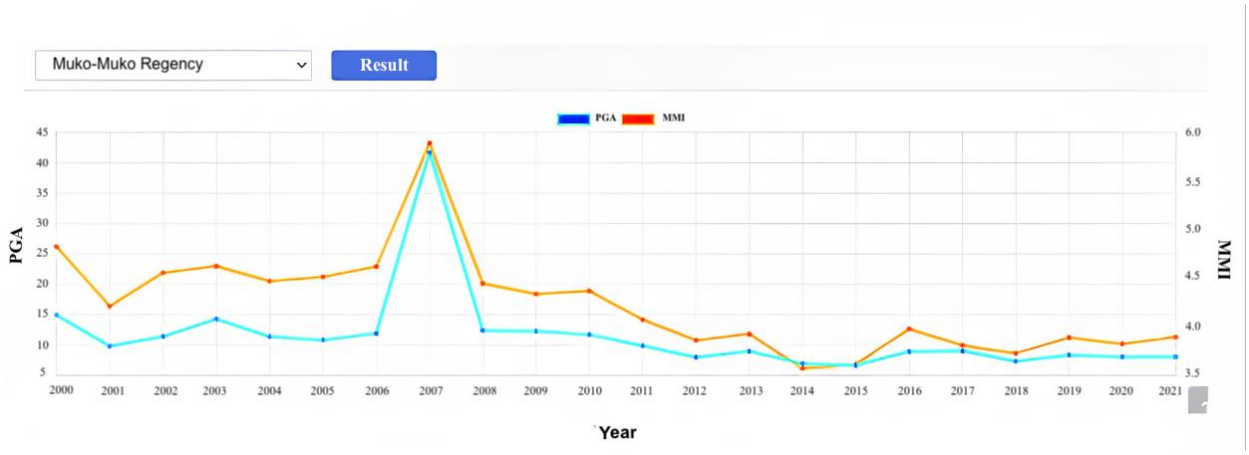


Fig. 6. Graph of the average PGA value and the average MMI value.

The largest average of PGA and the average value of MMI per year are in the Muko-Muko District. It can be seen in Fig. 6 that the Muko-Muko District in 2007 had an average PGA value of 41,619 cm/sec^2 and an average MMI value of 6. As seen from the graph (Fig. 6), Muko-Muko District in 2007 had an average value. The average PGA and the highest average MMI values compared to other years are seen from the peak, or highest point. The results of earthquake data clustering and earthquake intensity classification can be combined. This combination uses clustering results and the consequences of earthquake intensity (MMI) ≥ 6, which can produce new information. For example, in Table 5, Muko-Muko District is an area that experienced moderate damage in each cluster.

Table 5
Classification of MMI.

Cluster	Data	Description
C1 (Moderate Damage)	187	All districts in Bengkulu Province experienced moderate levels of damage.
C2 (Slight Damage)	31	Only Muko-Muko Regency did not experience light damage.
C3 (Slight Damage)	19	Central Bengkulu Regency, Kepahiang Regency, Rejang Lebong Regency, North Bengkulu Regency, South Bengkulu Regency, and Seluma Regency slight levels of damage. Bengkulu City, Lebong Regency, Kaur Regency, and Muko-Muko Regency moderate levels of damage.
C4 (Moderate Damage)	269	All districts in Bengkulu province experienced moderate levels of damage.
C5 (Moderate Damage)	467	Bengkulu City, Central Bengkulu Regency, Kepahiang Regency, and Seluma Regency slight levels of damage. Lebong Regency, Kaur Regency, Muko-Muko Regency, Rejang Lebong Regency, North Bengkulu Regency, and South Bengkulu Regency moderate levels of damage.
C6 (Moderate Damage)	2016	Bengkulu City, Central Bengkulu Regency, South Bengkulu Regency, Seluma Regency, and Kaur Regency slight levels of damage. Kepahiang Regency, Lebong Regency, Kaur Regency, Muko-Muko Regency, Rejang Lebong Regency, and North Bengkulu Regency moderate levels of damage.

In general, for a huge earthquake in 2007, the maximum possible damage in the MMI scale that districts in Bengkulu City could undergo is more than VIII. A recent study conducted by Mase et al. [55], predicts that the damage intensity level in the study area could be more serious, especially for areas having low soil resistance. The results of this study which present the average value of damage intensity level in the study area, are reasonable and consistent with several observations conducted by Hausler and Anderson [56] and Aydan et al. [57]. Overall, the framework implementation in this study has been successfully implemented and is applicable for implementation in other areas.

6. Conclusions

This research has succeeded in clustering earthquake data in Bengkulu Province from 2000-2021 using three methods, namely Elbow, CLARANS, and silhouette coefficient methods, and knowing the level of damage to buildings in each area based on the value of earthquake intensity (MMI). Several concluding remarks can be drawn:

- This study divided the cluster details into five clusters with the variation of epicenters, focal depths, and magnitudes. Basically, the magnitude of earthquake in the study area is observed to vary from 3 to 8.5 that occurred within 2000 to 2021. It indicates that the level of damage intensity level could be varied with the magnitude and the distance of earthquake focal depth
- Cluster interpretation is 0.55 (feasible cluster). This combination uses clustering and earthquake intensity (MMI) ≥ 6 is C1 moderate damage, C2 slight damage, C3 slight damage, C4 moderate damage, C5 moderate damage, and C6 moderate damage.
- In general, Muko-Muko District is prone to experiencing moderate levels of damage. The activity of Mentawai Fault located near to the district could be the main threat to produce earthquake. Therefore, seismic hazard mitigation and deep understanding on the possible damage of earthquake in the study area should be achieved.

The results also suggested the local government should consider the possible seismic damage in the study area. The enforcement on the implementation of seismic design code and early warning system should be performed to minimize the possible damage in the study area.

Acknowledgments

We would like to thank you to Meteorology, Climate, and Geophysics Body (BMKG) of Indonesia for providing one of the main datasets in this research. This research is hard to be done without the supporting data.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflict of interest.

Authors contribution statement

AV, FPU: Conceptualization; IPH: Data curation; LZM: Formal analysis; IPH: Investigation; AV, FPU: Methodology; FPU: Project administration; IPH: Resources; AV, IPH: Software; AV: Supervision; FPU, IPH: Validation; IPH: Visualization; AV, FPU, IPH, LZM: Roles/Writing – original draft; AV, LZM: Writing – review & editing.

References

[1] Okuwaki R, Hirano S, Yagi Y, Shimizu K. Inchworm-like source evolution through a geometrically complex fault fueled persistent supershear rupture during the 2018 Palu Indonesia earthquake. Earth Planet Sci Lett 2020;547:116449. https://doi.org/10.1016/j.epsl.2020.116449.

[2] Barakbah AR, Harsono T, Sudarsono A. Automatic Cluster-oriented Seismicity Prediction Analysis of Earthquake Data Distribution in Indonesia. Int J Adv Sci Eng Inf Technol 2019;9:587–93. https://doi.org/10.18517/ijaseit.9.2.7269.

[3] Goda K, Mori N, Yasuda T, Prasetyo A, Muhammad A, Tsujio D. Cascading Geological Hazards and Risks of the 2018 Sulawesi Indonesia Earthquake and Sensitivity Analysis of Tsunami Inundation Simulations. Front Earth Sci 2019;7. https://doi.org/10.3389/feart.2019.00261.

[4] Cilia MG, Mooney WD, Nugroho C. Field Insights and Analysis of the 2018 Mw 7.5 Palu, Indonesia Earthquake, Tsunami and Landslides. Pure Appl Geophys 2021;178:4891–920. https://doi.org/10.1007/s00024-021-02852-6.

[5] Supendi P, Ramdhan M, Priyobudi, Sianipar D, Wibowo A, Gunawan MT, et al. Foreshock–mainshock–aftershock sequence analysis of the 14 January 2021 (Mw 6.2) Mamuju–Majene (West Sulawesi, Indonesia) earthquake. Earth, Planets Sp 2021;73:106. https://doi.org/10.1186/s40623-021-01436-x.

[6] Sasmi AT, Nugraha AD, Muzli M, Widiyantoro S, Zulfakriza Z, Wei S, et al. Hypocenter and Magnitude Analysis of Aftershocks of the 2018 Lombok, Indonesia, Earthquakes Using Local Seismographic Networks. Seismol Res Lett 2020;91:2152–62. https://doi.org/10.1785/0220190348.

[7] Farid M. Implementation of Seismic Hazard Mitigation on THE Basis of Ground Shear Strain Indicator for Spatial Plan OF Bengkulu City, Indonesia. Int J GEOMATE 2020;18. https://doi.org/10.21660/2020.69.24759.

[8] Mase LZ. Local seismic hazard map based on the response spectra of stiff and very dense soils in Bengkulu city, Indonesia. Geod Geodyn 2022;13:573–84. https://doi.org/10.1016/j.geog.2022.05.003.

[9] Awaluddin M, Yuwono BD, Sajagat MJ. Study of interseismic coupling rate of 2007 Mentawai earthquake using GNSS data, 2017, p. 040004. https://doi.org/10.1063/1.4987068.

[10] Misliniyati R, Mase LZ, Syahbana AJ, Soebowo E. Seismic hazard mitigation for Bengkulu Coastal area based on site class analysis. IOP Conf Ser Earth Environ Sci 2018;212:012004. https://doi.org/10.1088/1755-1315/212/1/012004.

[11] Mase LZ, Sugianto N, Refrizon. Seismic hazard microzonation of Bengkulu City, Indonesia. Geoenvironmental Disasters 2021;8:5. https://doi.org/10.1186/s40677-021-00178-y.

[12] Sugianto N, Refrizon, Mase LZ, Anggeraini S. The updated zonation of Seismic Vulnerability Index and ground shear strain at Bengkulu City, Indonesia, 2021, p. 040022. https://doi.org/10.1063/5.0037597.

[13] Refrizon, Mase LZ. Geophysical observation to several sites around Pulau Baai Port, Bengkulu city, Indonesia, 2021, p. 040024. https://doi.org/10.1063/5.0037643.

[14] Fang L, Li Z. Preface to the special issue of Artificial Intelligence in Seismology. Earthq Sci 2023;36:81–3.

[15] Zhao M, Xiao Z, Chen S, Fang L. DiTing: A large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology. Earthq Sci 2023;36:84–94. https://doi.org/10.1016/j.eqs.2022.01.022.

[16] Soori M, Arezoo B, Dastres R. Machine learning and artificial intelligence in CNC machine tools, A review. Sustain Manuf Serv Econ 2023;2:100009. https://doi.org/10.1016/j.smse.2023.100009.

[17] Abbas KA, Gharavi A, Hindi NA, Hassan M, Alhosin HY, Gholinezhad J, et al. Unsupervised machine learning technique for classifying production zones in unconventional reservoirs. Int J Intell Networks 2023;4:29–37. https://doi.org/10.1016/j.ijin.2022.11.007.

[18] Kim H-S. Geospatial data-driven assessment of earthquake-induced liquefaction impact mapping using classifier and cluster ensembles. Appl Soft Comput 2023;140:110266. https://doi.org/10.1016/j.asoc.2023.110266.

[19] Schubert E, Rousseeuw PJ. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Inf Syst 2021;101:101804. https://doi.org/10.1016/j.is.2021.101804.

[20] Shi C, Wei B, Wei S, Wang W, Liu H, Liu J. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. EURASIP J Wirel Commun Netw 2021;2021:31. https://doi.org/10.1186/s13638-021-01910-w.

[21] Wu Z, Li F, Zhu Y, Lu K, Wu M, Zhang C. A Filter Pruning Method of CNN Models Based on Feature Maps Clustering. Appl Sci 2022;12:4541. https://doi.org/10.3390/app12094541.

[22] Pawirodikromo W. The estimated PGA map of the Mw6.4 2006 Yogyakarta Indonesia earthquake, constructed from the Modified Mercalli intensity IMM. Bull New Zeal Soc Earthq Eng 2018;51:92–104. https://doi.org/10.5459/bnzsee.51.2.92-104.

[23] Mendoza RB, Ramos N, Dimalanta C. High-resolution peak ground acceleration modeling using geographic information systems: A case study of the potentially active Central Cebu Fault System, Philippines. J Asian Earth Sci X 2022;7:100097. https://doi.org/10.1016/j.jaesx.2022.100097.

[24] Lu J, Li W, Zhan W, Tie Y. Distribution and Mobility of Coseismic Landslides Triggered by the 2018 Hokkaido Earthquake in Japan. Remote Sens 2022;14:3957. https://doi.org/10.3390/rs14163957.

[25] BMKG, Skala MMI (Modified Mercalli Intensity), Badan Meteorologi, Klimatologi, dan Geofisika. https://www.bmkg.go.id/gempabumi/skala-mmi.bmkg. 2022.

[26] Priyambada FR, Nugraha AD, Supendi P. Hypocenter Determination using a Non-Linear Method in Bali, Lombok, and Nusa Tenggara Regions: Preliminary Result. J Phys Conf Ser 2022;2243:012008. https://doi.org/10.1088/1742-6596/2243/1/012008.

[27] Rosalia S, Widiyantoro S, Dian Nugraha A, Ash Shiddiqi H, Supendi P, Wandono. Hypocenter Determination Using a Non-Linear Method for Events in West Java, Indonesia: A Preliminary Result. IOP Conf Ser Earth Environ Sci 2017;62:012052. https://doi.org/10.1088/1755-1315/62/1/012052.

[28] Seydoux L, Balestriero R, Poli P, Hoop M de, Campillo M, Baraniuk R. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nat Commun 2020;11:3972. https://doi.org/10.1038/s41467-020-17841-x.

[29] Rini DS, Novianti P, Fransiska H. Internal Cluster Validation on Earthquake Data in the Province of Bengkulu. IOP Conf Ser Mater Sci Eng 2018;335:012048. https://doi.org/10.1088/1757-899X/335/1/012048.

[30] Karamitsos I, Albarhami S, Apostolopoulos C. Applying DevOps Practices of Continuous Automation for Machine Learning. Information 2020;11:363. https://doi.org/10.3390/info11070363.

[31] Flores V, Heras S, Julián V. A New Methodological Framework for Project Design to Analyse and Prevent Students from Dropping Out of Higher Education. Electronics 2022;11:2902. https://doi.org/10.3390/electronics11182902.

[32] Spruit M, Kais M, Menger V. Automated Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding. Futur Internet 2021;13:243. https://doi.org/10.3390/fi13100243.

[33] Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Comput Sci 2021;181:526–34. https://doi.org/10.1016/j.procs.2021.01.199.

[34] Liu C, Chang Z, Li J. Characteristics of ground cracks caused by the MS 6.4 Yangbi earthquake and the corresponding tectonic significance. Earthq Res Adv 2022;2:100142. https://doi.org/10.1016/j.eqrea.2022.100142.

[35] Abasova J, Tanuska P, Rydzi S. Big Data—Knowledge Discovery in Production Industry Data Storages—Implementation of Best Practices. Appl Sci 2021;11:7648. https://doi.org/10.3390/app11167648.

[36] Ayele WY. Adapting CRISP-DM for Idea Mining. Int J Adv Comput Sci Appl 2020;11. https://doi.org/10.14569/IJACSA.2020.0110603.

[37] Yavary A, Sajedi H. Solving Dynamic Vehicle Routing Problem with Pickup and Delivery by CLARITY Method. 2018 IEEE 22nd Int. Conf. Intell. Eng. Syst., IEEE; 2018, p. 000207–12. https://doi.org/10.1109/INES.2018.8523908.

[38] Katyayan A, Bokhare A, Gupta R, Kumari S, Pardeshi T. Analysis of Unsupervised Machine Learning Techniques for Customer Segmentation, 2022, p. 483–98. https://doi.org/10.1007/978-981-16-7996-4_35.

[39] Carbajal SG. Customer Segmentation through Path Reconstruction. Sensors 2021;21:2007. https://doi.org/10.3390/s21062007.

[40] Bhuvaneswari MS, Muneeswaran K. User Community Detection From Web Server Log Using Between User Similarity Metric. Int J Comput Intell Syst 2020;14:266. https://doi.org/10.2991/ijcis.d.201126.002.

[41] Murjaya J, Pramumijoyo S, Karnawati D, Daryono, Meilano I, Supendi P, et al. Earthquake risk assessment of the Opak and Merapi-Merbabu active faults to support mitigation program in Yogyakarta province and its vicinity. IOP Conf Ser Earth Environ Sci 2021;851:012001. https://doi.org/10.1088/1755-1315/851/1/012001.

[42] Hason MM, Hanoon AN, Abdulhameed AA. Particle swarm optimization technique-based prediction of peak ground acceleration of Iraq's tectonic regions. J King Saud Univ - Eng Sci 2023;35:463–73. https://doi.org/10.1016/j.jksues.2021.06.004.

[43] Gandomi M, Soltanpour M, Zolfaghari MR, Gandomi AH. Prediction of peak ground acceleration of Iran's tectonic regions using a hybrid soft computing technique. Geosci Front 2016;7:75–82. https://doi.org/10.1016/j.gsf.2014.10.004.

[44] Ahmed MA, Baharin H, N.E. P. Analysis of K-means, DBSCAN and OPTICS Cluster Algorithms on Al-Quran Verses. Int J Adv Comput Sci Appl 2020;11. https://doi.org/10.14569/IJACSA.2020.0110832.

[45] Anitha P, Patil MM. RFM model for customer purchase behavior using K-Means algorithm. J King Saud Univ - Comput Inf Sci 2022;34:1785–92. https://doi.org/10.1016/j.jksuci.2019.12.011.

[46] Elkhouly A, Andrew AM, Rahim HA, Abdulaziz N, Abdulmalek M, Mohd Yasin MN, et al. A Novel Unsupervised Spectral Clustering for Pure-Tone Audiograms towards Hearing Aid Filter Bank Design and Initial Configurations. Appl Sci 2021;12:298. https://doi.org/10.3390/app12010298.

[47] Ghasemi SH, Bahrami H, Akbari M. Classification of Seismic Vulnerability Based on Machine Learning Techniques for RC Frames. J Soft Comput Civ Eng 2020;4:13–21. https://doi.org/10.22115/scce.2020.223322.1186.

[48] Ford E, Maneparambil K, Neithalath N. Machine Learning on Microstructural Chemical Maps to Classify Component Phases in Cement Pastes. J Soft Comput Civ Eng 2021;5:1–20. https://doi.org/10.22115/scce.2021.302400.1357.

[49] Morales L, Aguilar J. An Automatic Merge Technique to Improve the Clustering Quality Performed by LAMDA. IEEE Access 2020;8:162917–44. https://doi.org/10.1109/ACCESS.2020.3021675.

[50] Karo IMK, Huda AF. Spatial clustering for determining rescue shelter of flood disaster in South Bandung using CLARANS Algorithm with Polygon Dissimilarity Function. 2016 12th Int. Conf. Math. Stat. Their Appl., IEEE; 2016, p. 70–5. https://doi.org/10.1109/ICMSA.2016.7954311.

[51] Vanfretti L, Arava VSN. Decision tree-based classification of multiple operating conditions for power system voltage stability assessment. Int J Electr Power Energy Syst 2020;123:106251. https://doi.org/10.1016/j.ijepes.2020.106251.

[52] Jiao L, Yang H, Liu Z, Pan Q. Interpretable fuzzy clustering using unsupervised fuzzy decision trees. Inf Sci (Ny) 2022;611:540–63. https://doi.org/10.1016/j.ins.2022.08.077.

[53] Si D, Hu W, Deng Z, Xu Y. Fair hierarchical clustering of substations based on Gini coefficient. Glob Energy Interconnect 2021;4:576–86. https://doi.org/10.1016/j.gloei.2022.01.009.

[54] Li Q, Wang X, Pei Q, Chen X, Lam K-Y. Consistency preserving database watermarking algorithm for decision trees. Digit Commun Networks 2023. https://doi.org/10.1016/j.dcan.2022.12.015.

[55] Mase LZ, Wahyuni MS, Hardiansyah, Syahbana AJ. Prediction of Damage Intensity Level Distribution in Bengkulu City, During the Mw 8.6 Bengkulu-Mentawai Earthquake in 2007, Indonesia. Transp Infrastruct Geotechnol 2024;11:769–93. https://doi.org/10.1007/s40515-023-00306-1.

[56] Hausler E, Anderson A. Observation of the 12 and 13 September 2007 Earthquake, Sumatra, Indonesia. Build Chang Report, Denver, Color USA 2007.

[57] Aydan O, Imamura F, Suzuki T, Febrin I, Hakam A, Mera M, et al. A reconnaissance report on the Bengkulu Earthquake of September 12, 2007. Publ Jointly by Japan Soc Civ Eng Japan Assoc Earthq Eng 2007.